

AD-A253 036

## DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

It is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and reviewing the collection of information, Send comments regarding this burden estimate or any other aspect of this burden including this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Avenue, Washington, DC 20503, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 15 June 92	3. REPORT TYPE AND DATES COVERED Annual Progress Report 5/15/91-5/14/92
4. TITLE AND SUBTITLE Pattern Analysis Based Models of Masking by Spatially Separated Sound Sources		5. FUNDING NUMBERS PE-61102F G - AFOSR 91-0289 PR-2313 TA-CS	
6. AUTHOR(S) Robert H. Gilkey			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Wright State University Department of Psychology Dayton, Ohio 45435		8. PERFORMING ORGANIZATION REPORT NUMBER AFOSR-TR-92-0650	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research Bolling Air Force Base District of Columbia 20332 Dr Tangney		10. SPONSORING/MONITORING AGENCY REPORT NUMBER DTIC JUL 22 1992	
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; Distribution unlimited.		12b. DISTRIBUTION CODE	

## 13. ABSTRACT (Maximum 200 words)

One goal of this program of research is to examine masking among spatially separated sound sources. The results indicate that masking release on the order of 8-20 dB can be observed in free-field masking situations when the signal and the masker are spatially separated by 90°. This magnitude of masking release is comparable to that observed in traditional binaural masking level difference experiments, where the stimuli are presented through headphones. However, while masking release observed in headphone studies is typically assumed to be based on interaural differences in phase or on interaural differences in intensity, we observed substantial masking release in the median plane where interaural differences are minimal. Our own headphone masking research is also questioning traditional models of binaural masking. Work is underway to develop a neural network based model of sound localization. The results of these studies will have implications for the development of virtual environments and auditory displays.

14. SUBJECT TERMS			15. NUMBER OF PAGES 11
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)  
Prescribed by ANSI Std Z39-18  
298-102

17 JUN 1992

92-19280



**Pattern Analysis Based Models of Masking  
by Spatially Separated Sound Sources  
AFOSR 91-0289  
Annual Progress Report  
May 15, 1991 to May 14, 1992**

## **I. SUMMARY**

Auditory masking is being investigated under free-field and headphone conditions. The results from the free-field studies show a substantial release from masking when the signal and masker sound sources are spatially separated by 90° in either the horizontal, median, or frontal planes. The headphone studies compare diotic and dichotic stimulus presentation. Strong correlations were found between diotic and dichotic performance in cases where traditional models of binaural masking would predict a weak correlation. Conversely, weak correlations were observed in cases where binaural models would predict strong correlations. Overall both the free-field and headphone results indicate that traditional models need to be modified or replaced in order to deal with the available data. Efforts are currently underway to implement a neural network based model of sound localization and free-field masking.

The technology necessary for the synthesis of realistic three-dimensional sound images through headphones is rapidly emerging. This technology will be used to provide realistic auditory information in virtual environments and to create auditory displays. In designing these new systems it is important to know how the various auditory warnings, signals, and communication channels will interact and how to spatially arrange these stimuli in order to provide the most effective information transmission. Thus, the empirical and theoretical results from these basic studies should provide useful tools for system designers.

## **II. RESEARCH OBJECTIVES**

In the laboratory, the subject's task is frequently structured such that only a single sound source is present. In contrast, in most real world listening situations several sound sources are present. Often the observer must monitor one or more sources while ignoring others. Typically, the sound sources do not share a single location, but rather are distributed across space. One goal of this project is to examine the patterns of interference (masking) that are produced among multiple spatially separated sound sources. A second goal of this research is to develop a model of sound localization and of masking among spatially separated sound sources.

## **III. STATUS OF THE RESEARCH**

Most of the work described here, and planned for the future, is being conducted in the Auditory Localization Facility of the Armstrong Laboratory at Wright Patterson Air Force Base. This facility houses a fourteen foot diameter geodesic sphere with 272 speakers mounted at the sphere's vertices. Additional studies are being performed in the Signal Detection Laboratory in the Department of Psychology at Wright state University. The work on campus is supported by the

cost sharing funds from Wright State University, as part of this project, and by a grant from the National Institutes of Health.

#### A. Research Conducted at the Armstrong Laboratory.

**Experiments.** As discussed below, much of our effort during Year 1 of this project has been focused on refining and upgrading the Auditory Localization Facility to allow us to properly implement our masking experiments. Nevertheless, three experimental investigations were conducted using the facility. The three studies are closely related and differ more in terms of procedural and apparatus sophistication than in terms of intellectual thrust. The goal of all three experiments has been to measure the masked threshold for a signal as a function of a spatial separation between the signal and a masker.

At the end of the summer of 1991 the apparatus was in condition to allow us to run a brief pilot experiment. The experiment was intended largely to test the experimental procedures and evaluate the performance of the apparatus. During this experiment the subject stood on a small platform such that his head was approximately in the middle of the sphere. No head restraint was used, but the subjects were encouraged to hold their heads still. The signal was a 1000-Hz pure tone and the masker was a 100-3000-Hz white gaussian noise. Masking release was measured as the decrease in threshold for conditions where the signal and masker were separated by 90° in azimuth or in elevation, as compared to conditions where the signal and the masker were both presented through the same speaker. Three subjects participated in this experiment. For this pure tone signal, the only case where a notable release from masking was observed was when the signal was presented from directly in front of the subject at 0° elevation and the masker was presented from the subject's side at 90° azimuth and 0° elevation. The observed masking release was on the order of 3-6 dB.

During the fall of 1991 a more extensive experiment employing a more spectrally complex signal was conducted. The signal was a 200-ms train of 10-μs pulses, with a 100-Hz repetition rate. The pulse train was band-pass filtered between .4 and 1.6 kHz, 1.5 and 6.0 kHz, or 5.0 and 20.0 kHz. The maskers were 250-ms bursts of white noise, which began 25 ms before the signal and ended 25 ms after the signal. The bandwidth of the maskers was 3 octaves and was adjusted for each signal such that the signal band fell in the center of the masker band. As in the summer experiment the subjects stood during the experiment and no head restraint was present. Several combinations of masker and speaker location were investigated in the horizontal, frontal, and median planes. Two subjects participated in this experiment. Substantial masking release was observed as the signal was moved away from the masker in each of the three planes. In the horizontal plane, masking release as great as 18-19 dB was observed when the signal and masker were separated by 90°. Masking release as great as 12-13 dB was observed with a 100° separation in the frontal plane. In the median plane, reductions in masking on the order of 10-11 dB were observed with a 90° separation, for some combinations of subjects and conditions. In general, the high frequency signal showed the largest release from masking. This was particularly true in the median plane where little or no release from masking was observed for either the low or middle

or	<input checked="checked" type="checkbox"/>
Codes	
Dist	
and/or	
Special	

A-1

RECEIVED DIRECTED 3

frequency signal.

By the spring of 1992 a chair had been installed, which allowed subjects to be seated during the experimental session. A bite bar was used to limit their head movements. The conditions in the spring experiment were similar to those in the fall experiment. The signal was a 200-ms train of 25- $\mu$ s pulses, with a 100-Hz repetition rate. The signal was filtered from .35-.99 kHz, 1.7-4.8 kHz, or 5.0-14.1 kHz. The bandwidth of the noise maskers was 2.5 octaves and was adjusted for each signal such that the signal band fell in the center of the masker band. The masker was either continuous or was 200 ms in duration and gated simultaneously with the signal. The release from masking as the signal was moved away from the masker in space was examined for several combinations of masker location and signal location in the horizontal plane and in the median plane. Three subjects participated in this experiment. Reductions in masking on the order of 20 dB were observed in the horizontal plane for separations of 135°. The form of the data can be seen more completely in Figure 1. Masking release of 8-9 dB was observed in the median plane with separations of 90°. The form of the data can be seen more completely in Figure 2. As in the fall experiment the greatest masking release was observed for the high frequency signal. This was particularly true in the median plane where no masking release was observed for either the low frequency or the middle frequency signal. The overall form of the data was quite similar for the continuous and gated masker conditions. However, thresholds were about 3 dB lower under the continuous masker condition.

Overall, these results indicate that masking release on the order of 8-20 dB can be observed in free-field masking situations when the signal and the masker are spatially separated. This magnitude of masking release is comparable to that observed in traditional binaural masking level difference experiments where the stimuli are presented through headphones. Potentially, free-field masking release observed in the horizontal plane, and perhaps in the frontal plane, can be explained by traditional models of binaural masking. However, the masking release predicted by these models depends on interaural differences in time or intensity. The masking release observed in our experiments within the median plane cannot readily be explained by such models. That is, to a first approximation, there are no interaural differences in time or intensity for stimuli presented within the median plane. The fact that greater masking release was observed for high frequency signals is compatible with current views that suggest that high frequency spectral alterations introduced by the pinnae play a significant role in sound localization. Our results are similar to those of Saberi et al. [J. Acoust. Soc. Am., 90: 1355-1370, 1991], but indicate that the masking release they observed in the median plane was probably determined by the high frequency portion of their waveforms.

Modeling free-field localization and masking. Traditional models of localization and binaural masking depend on interaural differences in time and interaural differences in intensity. While such models may be adequate to explain localization and masking within the horizontal plane, they are inadequate to explain localization and masking phenomena observed in the median plane, where no interaural differences exist. Current localization research suggests that location

dependent spectral variations introduced by the pinnae provide the basis for vertical localization. That is, as the elevation of the sound source changes, the distribution of energy in the high frequency region of the spectrum is altered. Therefore, vertical localization can be viewed as a complex pattern analysis task. Similarly, in a free-field masking task where both signal and masker sources are in the median plane we might expect that the spectral pattern on signal-plus-noise trials would be different than the spectral pattern on noise-alone trials and that this difference might form the basis for signal detection.

Artificial neural networks have had a great deal of success when applied to a number of pattern analysis tasks. Therefore, their application to these tasks seems appropriate. We are currently implementing neural networks to evaluate localization performance based on monaural spectral cues and the interaural cross correlation function. If successful, these modeling efforts will be extended to our free-field masking data.

Laboratory development. A number of modifications to the Auditory Localization Facility have been implemented during the past year to provide for more efficient data collection, higher fidelity sound production, and experiment specific enhancements.

Previously, a single 80386/33 personal computer had been used for controlling the geodesic sphere, for controlling the localization cue synthesizers (used to present localized sound images through headphones), for hardware and software development on both systems, and for demonstrations for laboratory visitors. Hence, all of these activities were limited by the availability of a single personal computer. Therefore, a 80486/33 based personal computer was purchased to be dedicated exclusively to controlling experiments in the geodesic sphere. Because on many days the sphere is in use eight or more hours a day, a 80386/40 based personal computer has also been purchased for program development. In addition, the Armstrong Laboratory has purchased a small lap-top personal computer which can be used for controlling demonstrations in the laboratory and elsewhere. As a result of these acquisitions there is now relatively good availability of these facilities. Software drivers have been developed on the 80486 for sound generation, for sound processing and production, for controlling the sphere, programmable attenuators and visual displays, and for recording subject responses.

Computer software has been developed to perform an adaptive, cued, two-alternative, forced-choice task. Interval marking lights and visual feedback are provided by LEDs on a visual display mounted directly in front of the subject at the same distance as the speakers. The signal level is varied adaptively to estimate the level that would produce 79.4% correct performance. The program allows selection of any two of 272 speakers for presentation of the signal and masker or both signal and masker can be presented from the same speaker.

Careful and consistent calibration of speakers is an important component of most localization and free-field masking research. During the past year we have taken a variety of steps toward our eventual goal of achieving an equivalent, spectrally-flat transfer characteristic for each speaker. That is, the aim is to be able to produce sounds such that the speakers are transparent, or at a minimum, provide no differential coloration to the sound. To this end, during the summer of

1991, student workers on this project glued and recaulked the baffles on the 272 speakers in the sphere. During this time, they checked the wiring and replaced damaged speakers. During our fall experimental sessions we discovered that the anechoic chamber, which houses the sphere, and the control room of the Auditory Localization Facility were powered by totally separate electrical services. This led to substantial electrical hum under most configurations and the Armstrong Laboratory installed new electrical outlets in the anechoic chamber from the same electrical service as the control room. This substantially reduced the observed hum. Additional ground loop problems that were discovered in the switching hardware for the sphere itself were corrected. Acoustic noise from the switching hardware was reduced by constructing a sound barrier between the switching hardware and the sphere.

In order to allow us to get equivalent calibration for each speaker, a rotating microphone stand is being developed which will, under computer control, turn the microphone to face each of the speakers. This will allow us to get comparable measurements from each speaker automatically. We intend to use these recordings to flatten and equalize the response of each speaker.

The Auditory Localization Facility was initially designed to allow subjects to perform in a localization experiment where they would turn and face the sound source. Therefore, the design allowed for a standing subject. For our experiments it is critical that the subject's head be stationary (i.e., a slight head movement can dramatically change the effective signal-to-noise ratio). After trying a number of alternatives it was determined that a stationary head could be best achieved for a seated subject using a bite-bar. This required the design of a chair, a bite-bar, and an extension to the stand (the original stand was designed for a standing subject and therefore left the head of a seated subject considerably below the center of the sphere).

A response box was built and interfaced to the computer. Because some of the experiments required continuous noise, it was necessary to have lights to mark the observation intervals. Thus, the LED display described above was constructed and installed. An intercom and video camera were installed to increase the efficiency of data collection and to provide increased safety for the subjects. Because the original configuration of the sphere had no speakers at 0°, 90°, 180°, or 270° horizontal azimuth, nor a speaker directly above the subjects head, Armstrong Laboratories ordered new struts, hubs, and baffles, which had to be specially manufactured. These parts and new speakers were installed by student workers on this project.

The Auditory Localization Facility is utilized by several scientists engaged in a wide variety of projects. By the fall of 1992 it was decided that a full-time facilities manager was needed in order to oversee the development and installation of enhancements to the facility, and to assure the smooth operation of all projects. With support from AFOSR, the Armstrong Laboratory hired a full-time engineer to serve as facilities manager in December 1992. The progress made during the last six months is due, in large part, to his efforts.

Attempts to model localization and masking performance are using artificial neural networks. These programs are computationally intensive. Currently, we are using our own DECstation 5000 and ten SUN SPARCstations. Because the SUN workstations are used in classes, they are only available to us from midnight to eight a.m. each day. Therefore, we have

recently ordered two additional DECstation 5000s, whose primary function will be the development, implementation, and evaluation of our neural net models.

B. Research conducted at Wright State University.

**Experiments.** The large masking level difference (MLD) observed between monaural and binaural tone-in-noise masking tasks has been used to suggest that quite different processing is employed under these conditions (e.g., energy detection vs. interaural time processing). However, when Gilkey et al. [J. Acoust. Soc. Am., 78: 1207-1219, 1985] examined the trial-by-trial responses of subjects, they found that the responses under the NOS0 (both noise and signal presented diotically) and NOS $\pi$  (noise diotic, but signal presented 180° out of phase interaurally) conditions were highly correlated. That is, although the signal level under the NOS $\pi$  condition was 10 to 15 dB lower than under the NOS0 condition (because of the MLD), individual reproducible noise-alone or signal-plus-noise waveforms that were likely to elicit a positive response (i.e., a report of signal present) under one condition were also likely to elicit a positive response under the other condition. Gilkey et al. used wideband reproducible noise samples as maskers. When Isabelle and Colburn [J. Acoust. Soc. Am., 89: 352-359, 1991] examined the responses of subjects to narrowband reproducible noise samples, they found correlations that were much weaker and often negative. They attributed the differences between their data and those of Gilkey et al. to the differences in the bandwidth of the masker. However, Gilkey [Paper presented at the Midwinter Meeting of the Association for Research in Otolaryngology, February, 1990] directly compared narrowband and wideband results for the same subjects and found highly significant correlations between NOS0 and NOS $\pi$  conditions with both wideband and narrowband maskers. The correlation between NOS0 and NOS $\pi$  responses has significant implications for models of both monaural and binaural performance. Lateralization based models of binaural masking are unable to predict these data. On the other hand, Gilkey et al. showed that for a model such as the Equalization-Cancellation (EC) model [N.I. Durlach, "Binaural signal detection: Equalization and Cancellation Theory," in Foundations of Modern Auditory Theory II, edited by J.V. Tobias (Academic Press, New York) pp. 371-462, 1972] the effective maskers under the NOS0 and NOS $\pi$  conditions are highly correlated. Thus, the observed correlation of responses between these conditions is not necessarily unexpected.

In order to evaluate more fully the predictions of the EC model, we have measured the responses of subjects to individual reproducible waveforms under conditions where the effective masker at the output of the EC device should be quite different from the masker under the NOS0 condition. Under the NuS $\pi$  condition (independent noises to the two ears, signal 180° out of phase interaurally) the EC device should subtract the stimuli arriving from the two ears, such that the effective masker at the output of the EC device is the difference between the two monaural maskers. Thus, the EC model predicts that NOS0 responses to either of the two "monaural" maskers should be only partially correlated with the NuS $\pi$  responses. This is exactly what we observed in the data of our subjects. However, when the responses to the two monaural waveforms were averaged and then compared to the NuS $\pi$  responses, a very high correlation was

observed. This result was not anticipated. That is, this result would suggest that the stimuli in the two monaural channels are processed separately and not combined until the decision stage. [This work was presented at the Boston University Binaural Conference, December, 1991.]

Because we were surprised by this apparent failure of the EC model, we next examined a condition where the actual masker under the NOS0 condition was the difference between the two maskers presented under the NuS $\pi$  condition. That is, the masker under the NOS0 condition was the predicted effective masker under the NuS $\pi$  condition. Rather than the strong correlation we expected between these two conditions, based on the EC model, only a weak relation was observed. Overall the pattern of results suggests two possibilities, either: 1) the NuS $\pi$  condition is not a true binaural condition, as suggested by Durlach et al. [J. Acoust. Soc. Am., 79: 1548-1557, 1986], or 2) the EC model is an inadequate model of binaural hearing. [This work was presented at the Boston University Binaural Conference, December, 1991.]

Laboratory Development. In February 1991, the Signal Detection Laboratory was moved from Central Institute for the Deaf in St. Louis to Wright State University in Dayton. A considerable portion of our efforts during the past year have been applied to the process of setting up the new laboratory. The results of that process are described in brief. Wright State has been remodeling the rooms that house the laboratory and has provided four individual IAC sound attenuating booths. Computer hardware and software necessary for experimental control have been installed and are operational. Necessary analog hardware that was not brought from CID has been purchased or built. The current setup has functionality equivalent to that of the laboratory at CID. Data collection began in the Fall quarter 1991. A DECstation 5000 that was brought from CID has been installed and connected to the campus Ethernet network. This allows 2 Xterminals and a large-screen Macintosh computer to connect to the DECstation, providing X services to multiple simultaneous users. These systems provide data analysis, graphics, and modeling capabilities to the laboratory.

#### IV. PUBLICATION ACTIVITY

##### Papers in preparation

Good, M.D., and Gilkey, R.H. "Masking between spatially separated sounds."

#### V. PARTICIPATING PROFESSIONALS

##### Robert H. Gilkey

De Anza College, Cupertino, CA

University of California, Berkeley, CA

Indiana University, Bloomington, IN

B.A.

1976

Psychology

Ph.D.

1981

Psychology

Dissertation title: "Molecular psychophysics and models of auditory signal detectability."



## VI. INTERACTIONS

### Invited papers and conference talks

Gilkey, R. H. (1991). Headphone and Free-field Studies of Binaural Masking. Presented at the Boston University Binaural Conference, December, Boston, MA.

Good, M.D., and Gilkey, R.H. (1992). "Masking between spatially separated sounds." To be presented at The 36th Annual Meeting of the Human Factors Society, Atlanta, GA.

### Figure Captions

**Figure 1.** Mean masked thresholds for three subjects are plotted as a function of signal location within the horizontal plane. The abscissa represents the horizontal location of the signal, with 0° being directly in front of the subject. Each function represents a different frequency region for the signal. The position of the masker is indicated by the darkened arrows. The left panel shows the case where the masker is located at 0° azimuth and 0° elevation. The right panel shows the case where the masker is located at 90° azimuth and 0° elevation.

**Figure 2.** Mean masked thresholds for three subjects are plotted as a function of signal location within the median plane. The abscissa represents the elevation of the signal, with 0° being directly in front of the subject. Each function represents a different frequency region for the signal. The position of the masker is indicated by the darkened arrows. The left panel shows the case where the masker is located at 0° azimuth and 0° elevation. The right panel shows the case where the masker is located at 0° azimuth and 90° elevation.

FIGURE 1

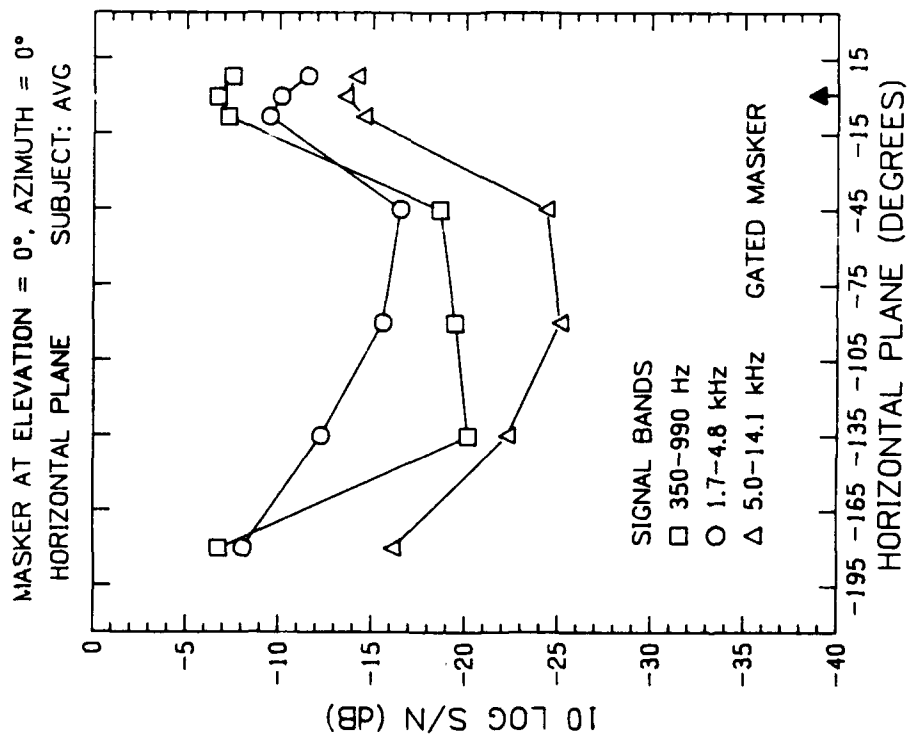
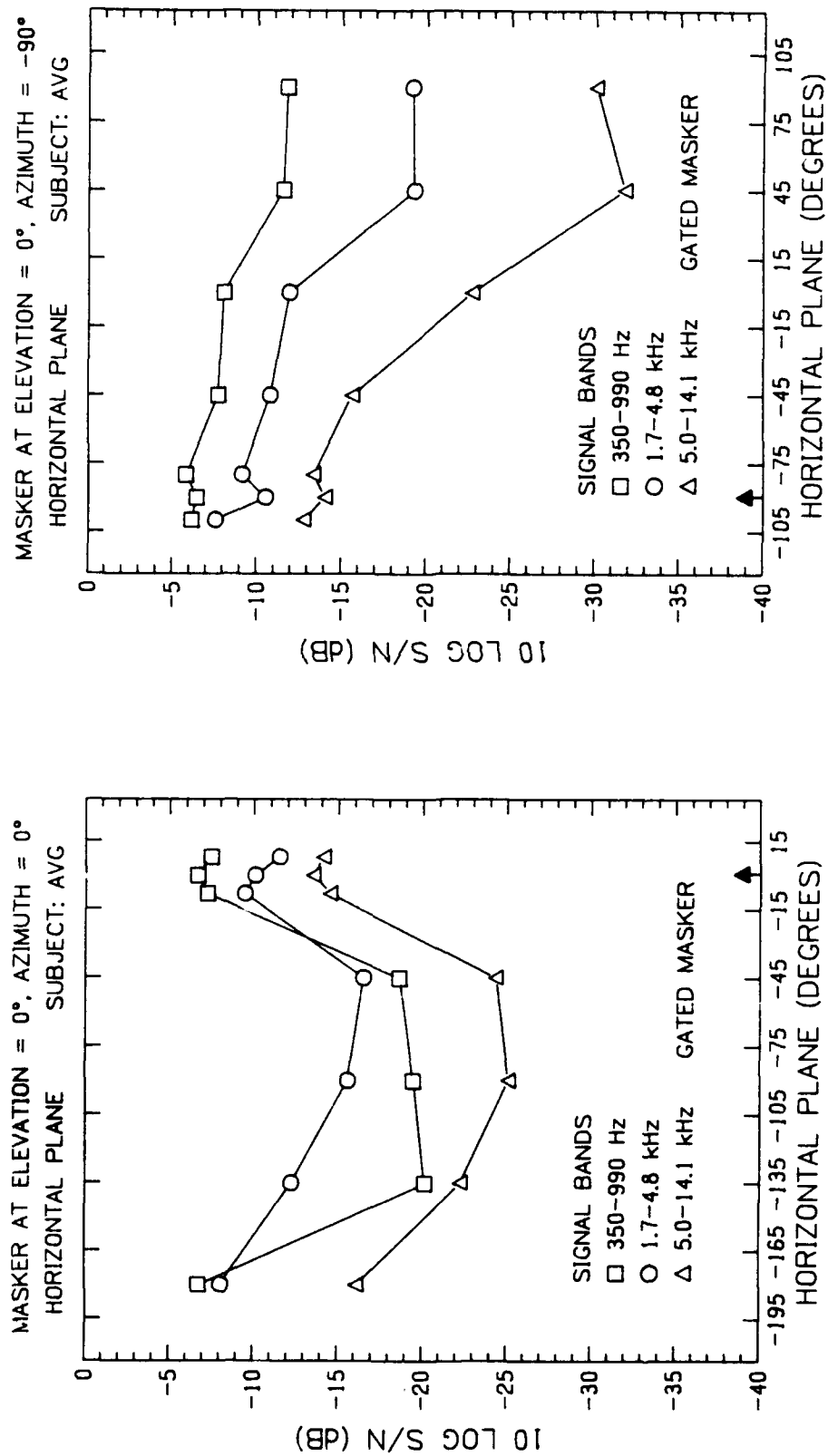


FIGURE 2

